REPORT

Alignment Analysis of Communication Arts Standards and Assessments

Missouri

Grades 3–8 and 11
2006 and 2007

Norman L. Webb

November 29, 2006

REPORT

Alignment Analysis of Communication Arts Standards and Assessments

Missouri
Grades 3–8 and 11
2006 and 2007

Norman L. Webb

November 29, 2006

Acknowledgements

Reviewers

John Fortier (Group Leader)	WI
Cindy Jacobson	WI
Joe Papenfuss	WI
Elizabeth VanderSchaaf	MN
Dana Humphrey	MO
Gwendolyn Cooke (Dr.)	MO
Debbie Jameson	MO
Melia Franklin	MO

Table of Contents

Executive Summary

At a three-day Alignment Institute held September 20-22, 2006, in Columbia, Missouri, eight reviewers analyzed the Communication Arts Standards and Assessments for 2006 and 2007 for grades 3–8 and 11. Reviewers included English language arts (ELA) content experts, district English language arts supervisors, and ELA teachers. Four of the reviewers were from Missouri and four were from other states. All reviewers analyzed the assessment for each of the two years, 2006 and 2007. The assessments only varied for Session 1, the first passage and six questions. Session 2 at each grade was the same for both years. The analysis weighted the items by the possible points that students could attain for each item. Each assessment had seven or eight items worth two points. The assessments for grades 3, 7, and 11 each had one writing prompt worth up to four points.

The alignment between the Communication Arts standards for reading and writing and the assessments for grades 3–8 and 11 was found to be reasonable. The alignment also was essentially the same for each year, 2006 and 2007. At grades 3, 5, 7, and 11, the only alignment weakness found was in the Range-of-Knowledge Correspondence and Balance of Representation criteria for the reading standard. For the other grades, only a Balance weakness was found. Since only one or two items would need to be replaced to reach an acceptable level on Range, this was not considered a critical alignment issue. The Balance issue was the result of the fact that most of the items for each grade were being coded to one of two grade-level expectations—text elements and features that included main idea and details. Because the other alignment criteria were generally met, Balance was not considered as critical an alignment issue. As long as there are an adequate number of items as a basis for making reliable judgments about students, the items have an appropriate level of complexity, and there is sufficient breadth, then it is not detrimental that some grade-level expectations are given more emphasis. In order to balance for the distribution of items among the grade-level expectations, from 4 to 10 items would need to be replaced by items that target less emphasized grade-level expectations. The Listening and Speaking Standard and the Information Literacy Standard purposefully were not included on the assessments. However, reviewers did code two multi-point items as corresponding to information literacy on the grade 8 2006 assessment. Overall, the alignment between the Reading and Writing Standards and the assessments for grades 3–8 and 11 was found to be reasonable, with only minor changes needed to attain full alignment.

Alignment Analysis of Communication Arts Standards and Assessments

Missouri
Grades 3–8 and 11

Norman L. Webb

Introduction

The alignment of expectations for student learning with assessments for measuring students' attainment of these expectations is an essential attribute for an effective standards-based education system. Alignment is defined as the degree to which expectations and assessments are in agreement and serve in conjunction with one another to guide an education system toward students learning what they are expected to know and do. As such, alignment is a quality of the relationship between expectations and assessments and not an attribute of any one of these two system components. Alignment describes the match between expectations and an assessment that can be legitimately improved by changing either student expectations or the assessment. As a relationship between two or more system components, alignment is determined by using the multiple criteria described in detail in a National Institute for Science Education (NISE) research monograph, *Criteria for Alignment of Expectations and Assessments in Mathematics and Science Education* (Webb, 1997).

A three-day Alignment Analysis Institute was conducted September 20-22, 2006, in Columbia, Missouri. Eight reviewers, including English language arts (ELA) content experts, district English language arts supervisors, and ELA teachers analyzed the agreement between the state's communication arts standards and the 2006 and 2007 assessments for grades 3–8 and 11. Four of the reviewers were from Missouri and four were from other states. All reviewers analyzed the assessment for each year, 2006 and 2007. The assessments only varied in Session 1, the first passage and six questions. Session 2 at each grade was the same for both years.

The State of Missouri uses the terminology of *standards* and *grade-level expectations* in its communication arts content expectations. Standards were the broad content requirements across all grades. Two communication arts standards were the main focus of this analysis—reading and writing. Two other communication arts standards— listening/speaking and information literacy—were included in the analysis, but these standards were not targeted by the assessments. Grade-level expectations (sometimes referred to as objectives) specified in greater detail under a standard what students are to know and do. The grade-level expectations were clustered under an intermediary level of expectations defined for the purpose of this analysis as goals. For example, the reading standard was divided into three goals (e.g., 1. Reading process strategies; 2. Comprehend, evaluate, and analyze fiction; and 3. Comprehend, evaluate, and analyze nonfiction). The goals were specified by two to seven grade-level expectations (e.g., Apply decoding strategies to "problem-solve" unknown words when reading). Some grade-level

expectations consisted of the same statement across grades. Data for this analysis were entered for the grade-level expectations and reported out at the standards level.

As part of the alignment institute, reviewers were trained to identify the depth of knowledge of the grade-level expectations and assessment items. This training included reviewing the definitions of the four depth-of-knowledge (DOK) levels and reviewing examples of each. Then the reviewers participated in 1) a consensus process to determine the depth-of-knowledge levels of the grade-level expectations and 2) individual analyses of the assessment items. Following individual analyses of the items, reviewers participated in a debriefing discussion in which they assessed the degree to which they had coded particular items or types of content to the grade-level expectations.

To derive the results from the analysis, the reviewers' responses are averaged. Any variance among reviewers is considered legitimate, with the true depth-of-knowledge level for the item falling somewhere between the two or more assigned values. Such variation could signify a lack of clarity in how the standards and grade-level expectations were written, the robustness of an item that can legitimately correspond to more than one grade-level expectation, and/or a depth of knowledge that falls between two of the four defined levels. Reviewers were allowed to identify one assessment item as corresponding to up to three grade-level expectations—one primary hit (grade-level expectation) and up to two secondary hits. However, reviewers could only code one depth-of-knowledge level to each assessment item, even if the item corresponded to more than one grade-level expectation.

Reviewers were instructed to focus primarily on the alignment between the state standards and assessments. However, reviewers were encouraged to offer their opinions on the quality of the standards, or of the assessment activities/items, by writing a note about the item. Reviewers could also indicate whether there was a Source-of-Challenge issue with the item—i.e., a problem with the item that might cause the student who knows the material to give a wrong answer, or enable someone who does not have the knowledge being tested to answer the item correctly.

The results produced from the institute pertain only to the issue of alignment between the Missouri state standards and the state assessment instruments. Note that this alignment analysis does not serve as external verification of the general quality of the state's standards or assessments. Rather, only the degree of alignment is discussed in the results. For these results, the means of the reviewers' coding were used to determine whether the alignment criteria were met. When reviewers did vary in their judgments, the means lessened the error that might result from any one reviewer's finding. Standard deviations are reported in the tables provided in the Appendix, which give one indication of the variance among reviewers.

The present report describes the results of an alignment study of grade-level expectations and both the 2006 and 2007 operational tests in mathematics for grades 3–8 and 11 in Missouri. The study addressed specific criteria related to the content agreement between the state standards and grade-level assessments. Four criteria received major

attention: categorical concurrence, depth-of-knowledge consistency, range-of-knowledge correspondence, and balance of representation.

Alignment Criteria Used for This Analysis

This analysis judged the alignment between the standards and the assessments on the basis of four criteria. Information is also reported on the quality of items by identifying items with Sources-of-Challenge and other issues. For each alignment criterion, an acceptable level was defined by what would be required to assure that a student had met the standards.

Categorical Concurrence

An important aspect of alignment between standards and assessments is whether both address the same content categories. The categorical concurrence criterion provides a very general indication of alignment if both documents incorporate the same content. *The criterion of categorical concurrence between standards and assessment is met if the same or consistent categories of content appear in both documents.* This criterion was judged by determining whether the assessment included items measuring content from each standard. The analysis assumed that the assessment had to have at least six items for measuring content from a standard in order for an acceptable level of categorical concurrence to exist between the standard and the assessment. The number of items, six, is based on estimating the number of items that could produce a reasonably reliable subscale for estimating students' mastery of content on that subscale. Of course, many factors have to be considered in determining what a reasonable number is, including the reliability of the subscale, the mean score, and cutoff score for determining mastery. Using a procedure developed by Subkoviak (1988) and assuming that the cutoff score is the mean and that the reliability of one item is .1, it was estimated that six items would produce an agreement coefficient of at least .63. This indicates that about 63% of the group would be consistently classified as masters or nonmasters if two equivalent test administrations were employed. The agreement coefficient would increase if the cutoff score is increased to one standard deviation from the mean to .77 and, with a cutoff score of 1.5 standard deviations from the mean, to .88. Usually states do not report student results by standards or require students to achieve a specified cutoff score on subscales related to a standard. If a state did do this, then the state would seek a higher agreement coefficient than .63. Six items were assumed as a minimum for an assessment measuring content knowledge related to a standard, and as a basis for making some decisions about a student's knowledge of that standard. If the mean for six items is 3 and one standard deviation is one item, then a cutoff score set at 4 would produce an agreement coefficient of .77. Any fewer items with a mean of one-half of the items would require a cutoff that would only allow a student to miss one item. This would be a very stringent requirement, considering a reasonable standard error of measurement on the subscale.

Depth-of-Knowledge Consistency

Standards and assessments can be aligned not only on the category of content covered by each, but also on the basis of the complexity of knowledge required by each. *Depth-of-knowledge consistency between standards and assessment indicates alignment if what is elicited from students on the assessment is as demanding cognitively as what students are expected to know and do as stated in the standards.* For consistency to exist between the assessment and the standard, as judged in this analysis, at least 50% of the items corresponding to a standard had to be at or above the level of knowledge of the standard: 50%, a conservative cutoff point, is based on the assumption that a minimal passing score for any one standard of 50% or higher would require the student to successfully answer at least some items at or above the depth-of-knowledge level of the corresponding standard. For example, assume an assessment included six items related to one standard and students were required to answer correctly four of those items to be judged proficient—i.e., 67% of the items. If three, 50%, of the six items were at or above the depth-of-knowledge level of the corresponding standards, then for a student to achieve a proficient score would require the student to answer correctly at least one item at or above the depth-of-knowledge level of one standard. Some leeway was used in this analysis on this criterion. If a standard had between 40% and 50% of items at or above the depth-of-knowledge levels of the standards, then it was reported that the criterion was "weakly" met.

Interpreting and assigning depth-of-knowledge levels to both grade-level expectations within standards and assessment items are essential requirements of the alignment analysis. In communication arts, four DOK levels were used to judge both reading and writing objectives and assessment tasks. The reading levels are based on Valencia and Wixson (2000, pp. 909-935). The writing levels were developed by Marshá Horton, Sharon O'Neal, and Phoebe Winter. First, descriptions are for reading and then for writing:

Reading Level 1. Level 1 requires students to receive or recite facts, or to use simple skills or abilities. Oral reading that does not include analysis of the text, as well as basic comprehension of a text, is included. Items require only a shallow understanding of the text presented and often consist of verbatim recall from text, slight paraphrasing of specific details from the text, or simple understanding of a single word or phrase. Some examples that represent, but do not constitute all of, Level 1 performance are:

- Support ideas by reference to verbatim or only slightly paraphrased details from the text.
- Use a dictionary to find the meanings of words.
- Recognize figurative language in a reading passage.

Reading Level 2. Level 2 includes the engagement of some mental processing beyond recalling or reproducing a response; it requires both comprehension and subsequent processing of text or portions of text. Inter-sentence analysis of inference is required. Some important concepts are covered, but not in a complex way. Standards and items at

this level may include words such as summarize, interpret, infer, classify, organize, collect, display, compare, and determine whether fact or opinion. Literal main ideas are stressed. A Level 2 assessment item may require students to apply skills and concepts that are covered in Level 1. However, items require closer understanding of text, possibly through the item's paraphrasing of both the question and the answer. Some examples that represent, but do not constitute all of, Level 2 performance are:

- Use context cues to identify the meaning of unfamiliar words, phrases, and expressions that could otherwise have multiple meanings.
- Predict a logical outcome based on information in a reading selection.
- Identify and summarize the major events in a narrative.

Reading Level 3. Deep knowledge becomes a greater focus at Level 3. Students are encouraged to go beyond the text; however, they are still required to show understanding of the ideas in the text. Students may be encouraged to explain, generalize, or connect ideas. Standards and items at Level 3 involve reasoning and planning. Students must be able to support their thinking. Items may involve abstract theme identification, inference across an entire passage, or students' application of prior knowledge. Items may also involve more superficial connections between texts. Some examples that represent, but do not constitute all of, Level 3 performance are:

- Explain or recognize how the author's purpose affects the interpretation of a reading selection.
- Summarize information from multiple sources to address a specific topic.
- Analyze and describe the characteristics of various types of literature.

Reading Level 4. Higher-order thinking is central and knowledge is deep at Level 4. The standard or assessment item at this level will probably be an extended activity, with extended time provided for completing it. The extended time period is not a distinguishing factor if the required work is only repetitive and does not require the application of significant conceptual understanding and higher-order thinking. Students take information from at least one passage of a text and are asked to apply this information to a new task. They may also be asked to develop hypotheses and perform complex analyses of the connections among texts. Some examples that represent, but do not constitute all of, Level 4 performance are:

- Analyze and synthesize information from multiple sources.
- Examine and explain alternative perspectives across a variety of sources.
- Describe and illustrate how common themes are found across texts from different cultures.

Writing Level 1. Level 1 requires the student to write or recite simple facts. The focus of this writing or recitation is not on complex synthesis or analysis, but on basic ideas. The students are asked to list ideas or words, as in a brainstorming activity, prior to written composition; are engaged in a simple spelling or vocabulary assessment; or they are asked to write simple sentences. Students are expected to write, speak, and edit using the conventions of Standard English. This includes using appropriate grammar, punctuation, capitalization, and spelling. Students demonstrate a basic understanding and appropriate use of reference materials such as a dictionary, thesaurus, or Web site. Some examples that represent, but do not constitute all of, Writing Level 1 performance are:

- Use punctuation marks correctly.
- Identify Standard English grammatical structures, including the correct use of verb tenses.

Writing Level 2. Level 2 requires some mental processing. At this level, students are engaged in first-draft writing, or in brief extemporaneous speaking for a limited number of purposes and audiences. Students are expected to begin connecting ideas, using a simple organizational structure. For example, students may be engaged in note-taking, outlining, or simple summaries. Text may be limited to one paragraph. Some examples that represent, but do not constitute all of, Level 2 performance are:

- Construct or edit compound or complex sentences, with attention to correct use of phrases and clauses.
- Use simple organizational strategies to structure written work.
- Write summaries that contain the main idea of the reading selection and pertinent details.

Writing Level 3. Level 3 requires some higher-level mental processing. Students are engaged in developing compositions that include multiple paragraphs. These compositions may include complex sentence structure and may demonstrate some synthesis and analysis. Students show awareness of their audience and purpose through focus, organization, and the use of appropriate compositional elements. The use of appropriate compositional elements includes such things as addressing chronological order in a narrative, or including supporting facts and details in an informational report. At this stage, students are engaged in editing and revising to improve the quality of the composition. Some examples that represent, but do not constitute all of, Level 3 performance are:

- Support ideas with details and examples.
- Use voice appropriate to the purpose and audience.
- Edit writing to produce a logical progression of ideas.

Writing Level 4. Higher-level thinking is central to Level 4. The standard at this level is a multi-paragraph composition that demonstrates the ability to synthesize and analyze complex ideas or themes. There is evidence of a deep awareness of purpose and audience. For example, informational papers include hypotheses and supporting evidence. Students

are expected to create compositions that demonstrate a distinctive voice and that stimulate the reader or listener to consider new perspectives on the ideas and themes addressed. An example that represents, but does not constitute all of, Level 4 performance is:

- Write an analysis of two selections, identifying the common theme and generating a purpose that is appropriate for both.

Range-of-Knowledge Correspondence

For standards and assessments to be aligned, the breadth of knowledge required on both should be comparable. *The range-of-knowledge criterion is used to judge whether a comparable span of knowledge expected of students by a standard is the same as, or corresponds to, the span of knowledge that students need in order to correctly answer the assessment items/activities.* The criterion for correspondence between span of knowledge for a standard and an assessment considers the number of objectives within the standard with one related assessment item/activity. Fifty percent of the grade-level expectations for a standard had to have at least one related assessment item in order for the alignment on this criterion to be judged acceptable. This level is based on the assumption that students' knowledge should be tested on content from over half of the domain of knowledge for a standard. This assumes that each grade-level expectation for a standard should be given equal weight. Depending on the balance in the distribution of items and the need to have a low number of items related to any one grade-level expectation, the requirement that assessment items need to be related to more than 50% of the grade-level expectations for an standard increases the likelihood that students will have to demonstrate knowledge on more than one grade-level expectation per standard to achieve a minimal passing score. As with the other criteria, a state may choose to make the acceptable level on this criterion more rigorous by requiring an assessment to include items related to a greater number of the grade-level expectations. However, any restriction on the number of items included on the test will place an upper limit on the number of grade-level expectations that can be assessed. Range-of-knowledge correspondence is more difficult to attain if the content expectations are partitioned among a greater number of standards and a large number of grade-level expectations. If 50% or more of the grade-level expectations for a standard had a corresponding assessment item, then the range-of-knowledge correspondence criterion was met. If between 40% and 50% of the grade-level expectations for a standard had a corresponding assessment item, the criterion was considered "weakly" met.

Balance of Representation

In addition to comparable depth and breadth of knowledge, aligned standards and assessments require that knowledge be distributed equally in both. The range-of-knowledge criterion only considers the number of grade-level expectations within a standard hit (a standard with a corresponding item); it does not take into consideration how the hits (or assessment items/activities) are distributed among these grade-level expectations. *The balance-of-representation criterion is used to indicate the degree to*

which one grade-level expectation is given more emphasis on the assessment than another. An index is used to judge the distribution of assessment items. This index only considers the grade-level expectations for a standard that have at least one hit—i.e., one related assessment item per grade-level expectation. The index is computed by considering the difference in the proportion of grade-level expectations and the proportion of hits assigned to a given grade-level expectation. An index value of 1 signifies perfect balance and is obtained if the hits (corresponding items) related to a standard are equally distributed among the grade-level expectations for the given standard. Index values that approach 0 signify that a large proportion of the hits are on only one or two of all of the objectives hit. Depending on the number of grade-level expectations and the number of hits, a unimodal distribution (most items related to one grade-level expectation and only one item related to each of the remaining grade-level expectations) has an index value of less than .5. A bimodal distribution has an index value of around .55 or .6. Index values of .7 or higher indicate that items/activities are distributed among all of the grade-level expectations at least to some degree (e.g., every grade-level expectation has at least two items) and is used as the acceptable level on this criterion. Index values between .6 and .7 indicate the balance-of-representation criterion has only been "weakly" met.

Source-of-Challenge Criterion

The Source-of-Challenge criterion is only used to identify items on which the major cognitive demand is inadvertently placed and is other than the targeted Communication Arts grade-level expectation, concept, or application. Cultural bias or specialized knowledge could be reasons for an item to have a Source-of-Challenge problem. Such item characteristics may result in some students not answering an assessment item, or answering an assessment item incorrectly, or at a lower level, even though they possess the understanding and skills being assessed.

Findings

Standards

Eight reviewers participated in the depth-of-knowledge (DOK) level consensus process for the standards and grade-level expectations for the Missouri communication arts standards. A summary of their deliberations is presented in Table 1. The complete group consensus values for each standard and grade-level expectation (GLE) can be found in Appendix A. The reviewers judged that the grade-level expectations increased in complexity over the grades. In the lower grades, the majority of the 32 to 34 expectations were judged to have a depth-of-knowledge level of 1 or 2 (e.g., Recite verbatim facts, oral reading, comprehension, and application of skills). With the increase in grades, the proportion of grade-level expectations reviewers judged to be at DOK Level 3 (e.g., Going beyond the text, generalizing, and reasoning) increased to over half of the grade-level expectations at grade 11. The level of complexity for communication arts was found to be moderate, with an increase in sophistication across the grades.

Table 1

Percent of Objectives by Depth-of-Knowledge (DOK) Levels for Grades 3–8 and 11, Missouri Alignment Analysis for Communication Arts, 2006 Study

Grade	Total Number of Objectives	DOK Level	# of objs by Level	% within std by Level
3	33	1	12	36
		2	14	42
		3	7	21
4	32	1	9	28
		2	15	46
		3	8	25
5	34	1	7	20
		2	17	50
		3	10	29
6	34	1	6	17
		2	16	47
		3	12	35
7	34	1	6	17
		2	13	38
		3	15	44
8	34	1	6	17
		2	11	32
		3	17	50
11	34	1	6	17
		2	9	26
		3	18	52
		4	1	2

The reviewers were told that within each standard (e.g., *Reading*), the grade-level expectations were intended to fully span the content of the standard. For this reason, the reviewers only coded items to a standard if there were no grade-level expectations that the item appeared to target. Such items are considered to target a generic grade-level expectation. A large number of items coded to generic grade-level expectations may indicate ways in which a standard's content is neither fully spanned nor described by its grade-level expectations. This may also simply indicate that these items are not as precise as the grade-level expectations. Table 2 shows the items on each of the seven assessments that were coded to a generic grade-level expectation by more than one reviewer for years 2006 and 2007. No item from Session 1 for any grade was coded to a generic grade-level expectation. The same items, all from Session 2 of the assessments, for each year were coded to generic grade-level expectations.

Table 2

Items Coded to Generic Objectives by More Than One Reviewer, Missouri Alignment
Analysis for Communication Arts, Grades 3–8 and 11, 2006 and 2007 Studies

Grade	Assessment Item Number	Generic GLE (Number of Reviewers)
3	7	W2 (3)
3	30	W2 (7)
3	40	W2 (3)
4	22	W2 (7)
4	42	W2 (2)
4	50	W2 (3)
4	51	W2 (3)
4	52	W2 (2)
5	35	W2 (2)
5	49	W2 (3)
5	50	W2 (3)
5	51	W2 (3)
5	52	W2 (3)
6	18	W2 (5)
6	46	W2 (3)
6	47	W2 (4)
6	48	W2 (4)
6	50	W2 (3)
6	51	W2 (3)
6	52	W2 (3)
7	7	W2 (4)
7	36	W2 (6)
7	38	W2 (6)
7	52	W2 (3)
7	53	W2 (3)
7	54	W2 (3)
8	12	R3 (3)
8	19	W2 (3)
8	20	W2 (3)
8	21	W2 (3)
8	22	W2 (3)
8	38	W2 (7)
8	39	W2 (3)
8	41	W2 (6)
8	42	W2 (3)
11	25	R3 (3)
11	7	W2 (4)
11	27	W2 (8)

The use of generic grade-level expectations across the seven grades related to two or three common omissions in the standards. For most generic GLEs, the reviewer did not find a GLE that clearly related to organization—in particular, organization of paragraphs. For example, students were asked to decide where in a paragraph an additional sentence should be inserted, or the student was expected to choose which of four sentences would be the best topic sentence for a given paragraph. Reviewers also did not find clear GLEs related to punctuation, or had difficulty distinguishing among different GLEs that related to punctuation in some way. For example, grade 3 Item 30 (Session 2, Item 23) asked students to write a sentence using correct capitalization and punctuation, but the writing GLEs W.2.b.3 and W.2.c.3 were too specific to fit the precise situation. Two or more reviewers coded the greatest number of items to generic GLEs at grade 8. Reviewers' notes in Appendix C frequently describe their reasons for using the generic GLEs. Because the use of generic GLEs only related to one or two areas, it can be concluded that the GLEs are clear and reviewers were able to find GLEs that match most of the items.

Alignment of Curriculum Standards and Assessments

The Communication Arts assessments for grades 3–8 and 11 had from 56 to 59 items (Table 3). Each assessment had 7 or 8 constructed respond items worth two points. Three grades had a writing prompt, each worth 4 points. The assessments' point value ranged from 63 to 69 points. Each year of the assessment had the same point value.

Table 3
Number of Items and Point Value by Grade for Missouri Communication Arts Assessments, Grades 3–8 and 11, 2006 and 2007 Study

Grade Level	Number of Items	Number of Items by Point Value			Total Point Value
		1 pts	2 pts	4 pts	
3	56	47	8	1	67
4	55	47	8		63
5	55	47	8		63
6	55	47	8		63
7	59	51	7	1	69
8	59	51	8		67
11	59	51	7	1	69

The results of the analysis for each of the four alignment criteria are summarized in Table 4. More detailed data on each of the criteria are given in Appendix B for 2006 and in Appendix E for 2007 in the first three tables. In Table 4, "YES" indicates that an acceptable level was attained between the assessment and the standard on the criterion. "WEAK" indicates that the criterion was nearly met, within a margin that could simply be due to error in the system. "NO" indicates that the criterion was not met by a noticeable margin—10% over an acceptable level for Depth-of-Knowledge Consistency, 10% over an acceptable level for Range-of-Knowledge Correspondence, and .1 under an

index value of .7 for Balance of Representation. "NT" under Categorical Concurrence indicates that the standard was not tested on the assessment. If reviewers coded on the average less than two items to any standard, then the standard was considered not to be tested. For standards that were found not to be tested, "NA" (non-applicable) is entered for the other alignment criteria because there were insufficient data to make any judgments on these criteria.

Alignment Results for the 2006 Assessments

Overall, the alignment among the communication arts 2006 assessments and reading and writing standards was reasonable. For all grades, 3–8 and 11, the assessment had a sufficient number of items to make reliable judgments about students for the reading and writing standards. The set of items for each standard were judged to have an adequate level of complexity compared to the corresponding grade-level expectations. For three of the grades (4, 6, and 8), the items on the assessment addressed a sufficient proportion of the grade-level expectations. However, for the other four grades, the Range-of-Knowledge Correspondence criterion was only weakly met. This was not considered a major alignment issue because, at each grade, the Range could be acceptably met by replacing one item. For each of the seven grades, one or two of the reading grade-level expectations were overemphasized, R.2.c and R.3.c, resulting in the Balance of Representation criterion either being not met or only weakly met. If the other three alignment criteria are met, then overemphasizing one or two grade-level expectations is not considered as important an issue. However, for grades 3, 5, 7, and 11, the Range was also not fully met. This indicates that GLEs R2.c and R.3.c were over-emphasized, while not a sufficient number of the other grade-level expectations were addressed. None of the alignment issues for the reading and writing standards are considered major because they could be resolved by replacing one or two items at each grade. As expected, the listening and speaking standard was not assessed for any grade, even though one or two reviewers did code a few items to this standard. Reviewers did code three items, on the average, to the information literacy standard for grade 8. These items were at an appropriate DOK level, but there were too few to meet the Categorical Concurrence and the Range-of-Knowledge Correspondence criteria. The alignment for each grade is discussed in greater detail below.

Grade 3

The alignment of the 2006 communication arts assessment and the standards for grade 3 is reasonable for reading and writing (Table 4.1). However, only 6 of the 14 GLEs under reading had at least one corresponding item. This is one fewer than the 50% of the GLEs that need to have at least one item to meet the minimum acceptable level for Range. Four of the GLEs with hits had from 1 to 3 corresponding items, while GLE R.2.c.3 had 17 corresponding items. This resulted in an imbalance for the distribution of hits. The Range could reach the minimum level by replacing at least one item that is mapped to GLE R.2.c.3 with an item that targets a GLE not currently assessed. The balance could be improved by replacing about five of the items mapped to GLE R.2.c.3

with items that assess content related to other reading GLEs, except for R.3.c.3. No items corresponded to the Listening and Speaking standard.

Table 4
Summary of Acceptable Levels on Alignment Criteria for Communication Arts Grades 3–8 and 11 Standards and Assessments for Missouri Alignment Analysis

Table 4.1
Summary of Acceptable Levels on Alignment Criteria for Communication Arts Grade 3 Standards and Assessments for Missouri Alignment Analysis, 2006 Study

Grade 3	Alignment Criteria			
Standards	Categorical Concurrence	Depth-of-Knowledge Consistency	Range of Knowledge	Balance of Represent ation
R - Reading	YES	YES	WEAK	NO
W - Writing	YES	YES	YES	YES
L - Listening and Speaking	NT	NA	NA	NA

Grade 4

At grade 4, the alignment between the reading and writing standards and the assessment are reasonable (Table 4.2). The only alignment issue for these two standards is a weak balance for the reading standard. Reviewers coded 26 items to GLE R.2.c.4 and fewer than 8 to any other GLE. Both Range and Balance would be strengthened if 10 items that target GLE R.2.c.4 were replaced by items that measure grade 4 GLEs that do not have corresponding items, or only one or two. Reviewers coded no items to either Standard L (Listening and Speaking), or Standard I (Information Literacy)

Table 4.2
Summary of Acceptable Levels on Alignment Criteria for Communication Arts Grade 4, Standards and Assessments for Missouri Alignment Analysis, 2006 Study

Grade 4	Alignment Criteria			
Standards	Categorical Concurrence	Depth-of-Knowledge Consistency	Range of Knowledge	Balance of Represent ation
R – Reading	YES	YES	YES	WEAK
W – Writing	YES	YES	YES	YES
L - Listening and Speaking	NT	NA	NA	NA
I - Information Literacy	NT	NA	NA	NA

Grade 5

Similar to grade 3, the alignment of the 2006 communication arts assessment and the standards for grade 5 is reasonable for reading and writing (Table 4.3). However, less than 6 of the 14 GLEs under reading had at least one corresponding item. Two more

13

GLEs need to have at least one corresponding item in order for 50% of the GLEs to meet the minimum acceptable level for Range. Three of the five GLEs had either one or 5 corresponding items, while GLE R.3.c.5 had 20 corresponding items and GLE R.2.c.5 had 13 corresponding items. This resulted in a large imbalance for the distribution of hits. The minimum requirement for Range could be reached by replacing at least 2 items that are currently mapped to GLE R.3.c.5 with items that target GLEs not currently assessed. To address this Balance issue would require replacing an additional 8 items that map to GLE R.3.c.5 with items that map to other reading GLEs. Balance would not be considered as critical an alignment issue if Range was not also an issue. One reviewer coded one item to a GLE under Standard L (Listening and Speaking) and five reviewers coded one item (Item 47—Item 41, Session 2) to a GLE under Standard I (Information Literacy). That is, Standards L and I were not measured on the assessment.

Table 4.3
Summary of Acceptable Levels on Alignment Criteria for Communication Arts Grade 5, Standards and Assessments for Missouri Alignment Analysis, 2006 Study

Grade 5	*Alignment Criteria*			
Standards	*Categorical Concurrence*	*Depth-of-Knowledge Consistency*	*Range of Knowledge*	*Balance of Representation*
R – Reading	YES	YES	WEAK	NO
W – Writing	YES	YES	YES	YES
L - Listening and Speaking	NT	NA	NA	NA
I - Information Literacy	NT	NA	NA	NA

Grade 6

At grade 6, similar to grade 4, the alignment between the reading and writing standards and the assessment are reasonable (Table 4.4). The only alignment issue for these two standards is an imbalance for the reading standard. Reviewers coded 15 items to GLE R.2.c.6, one item to each of five GLEs, and 5 and 8 items to two other GLEs respectively. Balance would be strengthened if 7 items that target GLE R.2.c.6 were replaced by items that measure grade 6 GLEs that have one or no corresponding items. One reviewer coded one item to Standard L (Listening and Speaking) and six reviewers coded one item (Item 17—Item 11, Session 2) to GLE I.1.b.6 under Standard I (Information Literacy).

14

Table 4.4
Summary of Acceptable Levels on Alignment Criteria for Communication Arts Grade 6,
Standards and Assessments for Missouri Alignment Analysis, 2006 Study

Grade 6	Alignment Criteria			
Standards	*Categorical Concurrence*	*Depth-of-Knowledge Consistency*	*Range of Knowledge*	*Balance of Represent ation*
R - Reading	YES	YES	YES	NO
W - Writing	YES	YES	YES	YES
L - Listening and Speaking	NT	NA	NA	NA
I - Information Literacy	NT	NA	NA	NA

Grade 7

The alignment between the grade 7 reading and writing standards and the 2006 assessment is reasonable, but could be improved in ways similar to the other grades (Table 4.5). For grade 7, both the Range and Balance were found to be weak. On the average, reviewers only coded items to 6 of the 14 GLEs under reading. They coded 19 items to GLE R.3.c.7 and one or 2 items to three other GLEs. The Range could be improved by replacing at least 2 items that were mapped to GLE R.3.c.7 with items that target reading GLEs that are currently not assessed. The Balance would be improved by replacing 7 more items that mapped to GLE R.3.c.7 to reading GLEs with fewer than 3 corresponding items. One reviewer coded one item to Standard L (Listening and Speaking) and five reviewers coded one item (Item 1) to GLE I.1.a.7 under Standard I (Information Literacy). In addition, 4 other items were coded by only one reviewer each to GLEs under Standard I.

Table 4.5
Summary of Acceptable Levels on Alignment Criteria for Communication Arts Grade 7,
Standards and Assessments for Missouri Alignment Analysis, 2006 Study

Grade 7	Alignment Criteria			
Standards	*Categorical Concurrence*	*Depth-of-Knowledge Consistency*	*Range of Knowledge*	*Balance of Represent ation*
R – Reading	YES	YES	WEAK	WEAK
W – Writing	YES	YES	YES	YES
L - Listening and Speaking	NT	NA	NA	NA
I - Information Literacy	NT	NA	NA	NA

Grade 8

At grade 8, similar to grades 4 and 6, the alignment between the reading and writing standards and the assessment are reasonable (Table 4.6). The only alignment issue for these two standards is an imbalance for the reading standard. Reviewers coded

12 items each to GLE R.2.c.8 and GLE R.3.c.8. Only one to 3 items were coded to five other GLEs and 7 items to another GLE. Balance would be strengthened if 6 items, 3 from each of the two GLEs emphasized the most, were replaced by items that correspond to the GLEs with 3 or fewer corresponding items. However, attaining full Balance is not as critical for the reading standard because the other three alignment criteria were met. No reviewer coded any items to Standard L (Listening and Speaking). All eight reviewers coded Item 5 to a GLE under Standard I (Information Literacy) and three reviewers coded Item 27 to a GLE under Standard I. Because each of these items were worth two points each, the average number of hits for Standard I was three.

Table 4.6
Summary of Acceptable Levels on Alignment Criteria for Communication Arts Grade 8, Standards and Assessments for Missouri Alignment Analysis, 2006 Study

Grade 8	*Alignment Criteria*			
Standards	*Categorical Concurrence*	*Depth-of-Knowledge Consistency*	*Range of Knowledge*	*Balance of Represent ation*
R – Reading	YES	YES	YES	NO
W – Writing	YES	YES	YES	YES
L - Listening and Speaking	NT	NA	NA	NA
I - Information Literacy	NO	YES	NO	YES

Grade 11

The alignment between the grade 11 reading and writing standards and the 2006 assessment is reasonable, but could be improved in ways similar to those indicated for the other grades (Table 4.7). For grade 11, both the Range and Balance were found to be weak. On the average, reviewers only coded items to 6 of the 14 GLEs under reading, plus generic grade-level expectations (goals). They coded 18 items to GLE R.3.c.9-12 and 11 items to GLE R.2.c.9-12. Reviewers only found one or 2 items that corresponded to two or three other reading GLEs and 8 items that corresponded to GLE R.1.e.9-12. The Range could be improved by replacing at least 2 items that were mapped to GLE R.3.c.9-12 with items that target reading GLEs that are currently not assessed. The Balance would be improved by replacing 4 more items that mapped to GLE R.3.c.7 to reading GLEs with fewer than 3 corresponding items. No reviewers coded any items to Standard L (Listening and Speaking), or Standard I (Information Literacy).

Table 4.7
Summary of Acceptable Levels on Alignment Criteria for Communication Arts Grade 11,
Standards and Assessments for Missouri Alignment Analysis, 2006 Study

Grade 11	Alignment Criteria			
Standards	*Categorical Concurrence*	*Depth-of-Knowledge Consistency*	*Range of Knowledge*	*Balance of Represent ation*
R – Reading	YES	YES	WEAK	WEAK
W – Writing	YES	YES	YES	YES
L - Listening and Speaking	NT	NA	NA	NA
I - Information Literacy	NT	NA	NA	NA

Alignment Results for the 2007 Assessments

  The alignment results for the 2007 assessments of the Communication Arts standards are presented in Table 5. The 2007 assessments and the 2006 assessments only varied by the first six items, Session 1. Session 2 for each year was the same for each grade. As a consequence, the alignment results are nearly identical for both years of the assessments. Since the explanation on the alignment for the 2006 assessments given above applies for nearly all grades for the 2007 assessment, the explanation is not repeated for 2007. For the grade 6 assessment in 2007, the reading standard had for Balance of Representation a "weak" rather than a "no" finding. The Balance index varied so little for each year that there is no difference in the results, or needed changes to improve the Balance. At grade 7, reviewers coded on the average 2.5 items that corresponded to GLEs under Standard I (Information Literacy), just above the cutoff score for retaining information on the other alignment criteria. However, grade 8 Item 5 on the 2006 assessment that was coded to a GLE under Standard I was replaced on the 2007 assessment with an item that did not relate to Standard I. Thus, alignment criteria information is include for Standard I in Table 5.5 for grade 7, but not in Table 5.6 for grade 8. Finally, the 2007 assessment and the reading standard for grade 11 only weakly met an acceptable level for the Depth-of-Knowledge Consistency criterion. On the 2006 assessment, 51% of the 49 item points, on the average, that were mapped to the reading standard had a DOK level that was the same as or higher than the DOK level of the corresponding GLE. This just met the acceptable level for the DOK of 50% or higher. For the 2007 assessment, reviewers mapped 47% of the 48 item points, on the average, to GLEs under the reading standard that had the same or higher DOK level. Whereas Item 6 on the 2006 assessment was mapped to a reading GLE by six reviewers, Item 6 on the 2007 assessment was mapped to a reading GLE by four reviewers. The difference in the DOK level between the two years of assessments at grade 11 is very small. In both years and also for other grades, only the minimum acceptable level of depth of knowledge was attained. This Depth-of-Knowledge Consistency for both years can be improved by assuring that items replaced to improve Range and Balance have the appropriate DOK level.

Table 5
Summary of Acceptable Levels on Alignment Criteria for Communication Arts Grades 3–8 and 11 Standards and 2007Assessments for Missouri Alignment Analysis

Table 5.1
Summary of Acceptable Levels on Alignment Criteria for Communication Arts Grade 3 Standards and Assessments for Missouri Alignment Analysis, 2007 Study

Grade 3	*Alignment Criteria*			
Standards	*Categorical Concurrence*	*Depth-of-Knowledge Consistency*	*Range of Knowledge*	*Balance of Represent ation*
R – Reading	YES	YES	WEAK	NO
W – Writing	YES	YES	YES	YES
L - Listening and Speaking	NT	NA	NA	NA

Table 5.2
Summary of Acceptable Levels on Alignment Criteria for Communication Arts Grade 4 Standards and Assessments for Missouri Alignment Analysis, 2007 Study

Grade 4	*Alignment Criteria*			
Standards	*Categorical Concurrence*	*Depth-of-Knowledge Consistency*	*Range of Knowledge*	*Balance of Represent ation*
R - Reading	YES	YES	YES	WEAK
W - Writing	YES	YES	YES	YES
L - Listening and Speaking	NT	NA	NA	NA
I - Information Literacy	NT	NA	NA	NA

Table 5.3
Summary of Acceptable Levels on Alignment Criteria for Communication Arts Grade 5 Standards and Assessments for Missouri Alignment Analysis, 2007 Study

Grade 5	*Alignment Criteria*			
Standards	*Categorical Concurrence*	*Depth-of-Knowledge Consistency*	*Range of Knowledge*	*Balance of Represent ation*
R - Reading	YES	YES	WEAK	NO
W - Writing	YES	YES	YES	YES
L - Listening and Speaking	NT	NA	NA	NA
I - Information Literacy	NT	NA	NA	NA

Table 5.4

Summary of Acceptable Levels on Alignment Criteria for Communication Arts Grade 6 Standards and Assessments for Missouri Alignment Analysis, 2007 Study

Grade 6	Alignment Criteria			
Standards	Categorical Concurrence	Depth-of-Knowledge Consistency	Range of Knowledge	Balance of Represent ation
R - Reading	YES	YES	YES	WEAK
W - Writing	YES	YES	YES	YES
L - Listening and Speaking	NT	NA	NA	NA
I - Information Literacy	NT	NA	NA	NA

Table 5.5

Summary of Acceptable Levels on Alignment Criteria for Communication Arts Grade 7 Standards and Assessments for Missouri Alignment Analysis, 2007 Study

Grade 7	Alignment Criteria			
Standards	Categorical Concurrence	Depth-of-Knowledge Consistency	Range of Knowledge	Balance of Represent ation
R - Reading	YES	YES	WEAK	WEAK
W - Writing	YES	YES	YES	YES
L - Listening and Speaking	NT	NA	NA	NA
I - Information Literacy	NO	NO	NO	WEAK

Table 5.6

Summary of Acceptable Levels on Alignment Criteria for Communication Arts Grade 8 Standards and Assessments for Missouri Alignment Analysis, 2007 Study

Grade 8	Alignment Criteria			
Standards	Categorical Concurrence	Depth-of-Knowledge Consistency	Range of Knowledge	Balance of Represent ation
R - Reading	YES	YES	YES	NO
W - Writing	YES	YES	YES	YES
L - Listening and Speaking	NT	NA	NA	NA
I - Information Literacy	NT	NA	NA	NA

Table 5.7
Summary of Acceptable Levels on Alignment Criteria for Communication Arts Grade 11
Standards and Assessments for Missouri Alignment Analysis, 2007 Study

Grade 11	Alignment Criteria			
Standards	*Categorical Concurrence*	*Depth-of-Knowledge Consistency*	*Range of Knowledge*	*Balance of Represent ation*
R - Reading	YES	WEAK	WEAK	NO
W - Writing	YES	YES	YES	YES
L - Listening and Speaking	NT	NA	NA	NA
I - Information Literacy	NT	NA	NA	NA

Reviewers' Comments

Reviewers were instructed to document any Source-of-Challenge issue and to provide any other comments they may have. These comments can be found in Tables (grade).5 and (grade).7 in Appendix C (2006) and Appendix F (2007). After coding each grade-level assessment, reviewers also were asked to respond to five debriefing questions. All of the comments made by the reviewers are given in Appendix D (2006) and Appendix G (2007). The notes, in general, offer an opinion on the item or give an explanation of the reviewers' coding.

Reliability Among Reviewers

The overall intraclass correlation among the communication arts reviewers' assignment of DOK levels to items was very high for the eight reviewers for both 2006 and 2007 assessments (Table 6). An intraclass correlation value greater than 0.8 generally indicates a high level of agreement among the reviewers. All 14 intraclass correlations in assigning a DOK level to items were above .9. A pairwise comparison is used to determine the degree of reliability of reviewer coding at the grade-level expectation level and at the standard level. The standard pairwise comparison values are high, all above .88. One contributing factor to the high agreement in assigning items to standards is that in the communication arts there were mainly only the two standards, reading and writing. The GLE pairwise agreement was lower and ranged between .57 and .70. This is the general range in agreement in assigning items to grade-level expectations.

The communications arts group leader noted some explanations for the lack of higher agreement among reviewers in assigning items to GLEs. He noted that there were generally two areas where the GLEs had some overlap and made it difficult to come to precise agreement:

1. In each test, anywhere from three to seven items presented a number of sentences. The student was to decide which of three or four sentences was correctly written, correctly punctuated, and correct as to capitalization. The GLEs included one that

addressed punctuation, one that addressed capitalization, one that addressed proper sentence structure, and yet another that addressed the writing process, including editing. Some reviewers chose to assign an item to the GLE representing the error most frequently found in the sentences. Others chose to use the more general editing GLE. Still others (most appropriately) decided to assign the item to each of the GLEs necessary to make a proper selection. Frequently, the result was failure of a majority of the reviewers to agree on a GLE. If items were confined to addressing a single GLE, or if the writing mechanics GLEs were combined in a single objective, this kind of disagreement would not occur.

2. A number of items addressed content that was not clearly apparent in the GLEs. Most dealt with organization, particularly within paragraphs. For example, 1) The student is asked to decide where in a paragraph an additional sentence should be inserted; or 2) The student is asked to choose which of four sentences would be the best topic sentence for a given paragraph. Reviewers assigned these items either to the "generic," more general, GLE or tried to find a GLE into which the item might possibly fit. In most cases, they wrote notes on the alignment tool, explaining their choices. Not surprisingly, for these items, a majority of reviewers did not choose the same GLE. This problem might be resolved if test makers did not include items for which there is no appropriate GLE in a state's standards. If the skill involved is significant, the problem may suggest that there is a gap in the GLE document.

Table 6
Intraclass and Pairwise Comparisons, Missouri Alignment Analysis for Communication Arts, Grades 3–8 and 11, 2006 Study

Grade	Intraclass Correlation	Pairwise Comparison:	Pairwise: Objective	Pairwise: Standard
2006				
3	.94	.76	.69	.93
4	.93	.71	.69	.93
5	.94	.75	.67	.91
6	.94	.76	.57	.88
7	.95	.78	.58	.92
8	.96	.78	.68	.90
11	.95	.79	.70	.95
2007				
3	.94	.75	.68	.92
4	.95	.74	.70	.95
5	.96	.79	.69	.91
6	.93	.76	.58	.88
7	.94	.77	.58	.90
8	.96	.78	.65	.89
11	.95	.79	.68	.94

Summary

At a three-day Alignment Institute held September 20-22, 2006, in Columbia, Missouri, eight reviewers analyzed the Communication Arts Standards and Assessments for 2006 and 2007 for grades 3–8 and 11. Reviewers included English language arts (ELA) content experts, district English language arts supervisors, and ELA teachers. Four of the reviewers were from Missouri and four were from other states. All reviewers analyzed the assessment for each year, 2006 and 2007. The assessments only varied for Session 1, the first passage and six questions. Session 2 at each grade was the same for both years. The analysis weighted the items by the possible points that students could attain for each item. Each assessment had seven or eight items worth two points. The assessments for grades 3, 7, and 11 each had one writing prompt worth up to four points.

The alignment between the Communication Arts standards for reading and writing and the assessments for grades 3–8 and 11 was found to be reasonable. The alignment also was essentially the same for each year, 2006 and 2007. At grades 3, 5, 7, and 11, the only alignment weakness found was in Range-of-Knowledge Correspondence and Balance of Representation for the reading standard. For the other grades, only a weakness with Balance was found. Since only one or two items would need to be replaced to reach an acceptable level on Range, this was not considered as a critical alignment issue. The Balance issue was the result of most of the items for each grade being coded to one of two grade-level expectations—text elements and features, including main idea and details. Because the other alignment criteria were generally met, Balance was not considered as critical an alignment issue. As long as there are an adequate number of items to make reliable judgments about students, the items have an appropriate level of complexity, and there is sufficient breadth, then it is not detrimental that some grade-level expectations are given more emphasis. In order to balance for the distribution of items among the grade-level expectations, from 4 to 10 items would need to be replaced by items that target less emphasized grade-level expectations. The Listening and Speaking Standard and the Information Literacy Standard purposefully were not included on the assessments. However, reviewers did code two multi-point items as corresponding to information literacy on the grade 8 2006 assessment. Overall, the alignment between the reading and writing standards and the assessments for grades 3–8 and 11 were found to be reasonable, with only minor changes needed to attain full alignment.

References

Subkoviak, M. J. (1988). A practitioner's guide to computation and interpretation of reliability indices for mastery tests. *Journal of Educational Measurement, 25*(1), 47-55.

Valencia, S. W., & Wixson, K. K. (2000). Policy-oriented research on literary standards and assessment. In M. L. Kamil, P. B. Mosenthal, P. D. Pearson, & R. Barr (Eds.), *Handbook of reading research: Vol III.* Mahwah, NJ: Lawrence Erlbaum.

Webb, N. L. (1997). *Criteria for alignment of expectations and assessments in mathematics and science education.* Council of Chief State School Officers and National Institute for Science Education Research Monograph No. 6. Madison: University of Missouri, Missouri Center for Education Research.